

PREDICCIÓN DE PARTIDOS DE FÚTBOL

Proyecto Machine Learning

Gonzalo Velázquez Martínez

THE BRIDGE



ÍNDICE

1. Objetivo - Hipótesis.
2. Otros posibles usos.
3. Pequeño Análisis EDA.
4. Modelos utilizados.
5. Problemas encontrados.
6. Predicciones.
7. Posibles mejoras futuras.
8. Conclusiones.



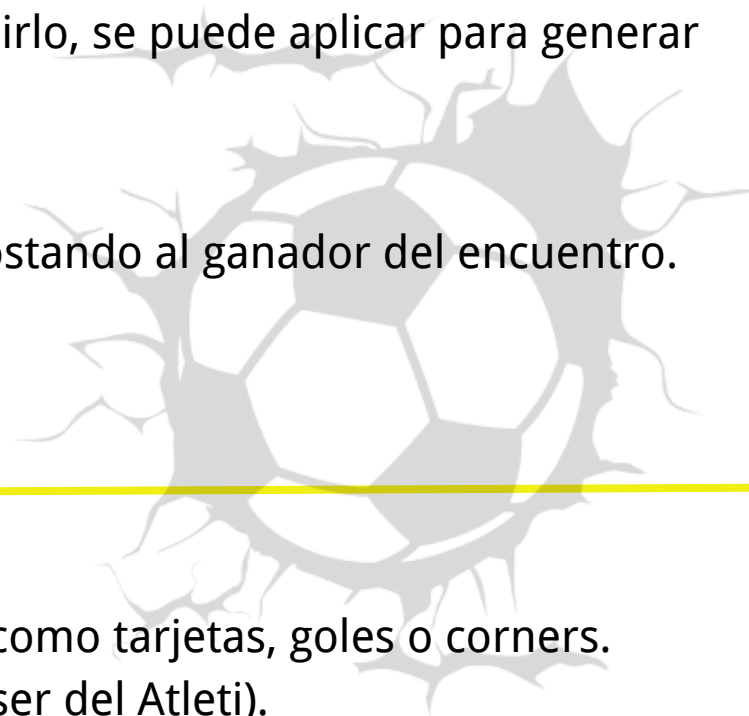
OBJETIVO - HIPOTESIS

El principal objetivo del proyecto es la clasificación de partidos de fútbol con el fin de predecir el ganador de un encuentro. De conseguirlo, se puede aplicar para generar beneficios en:

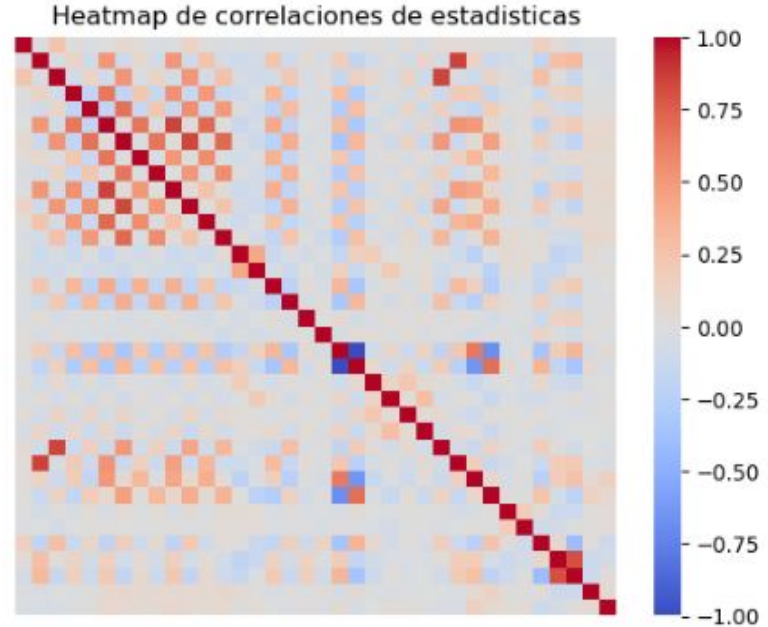
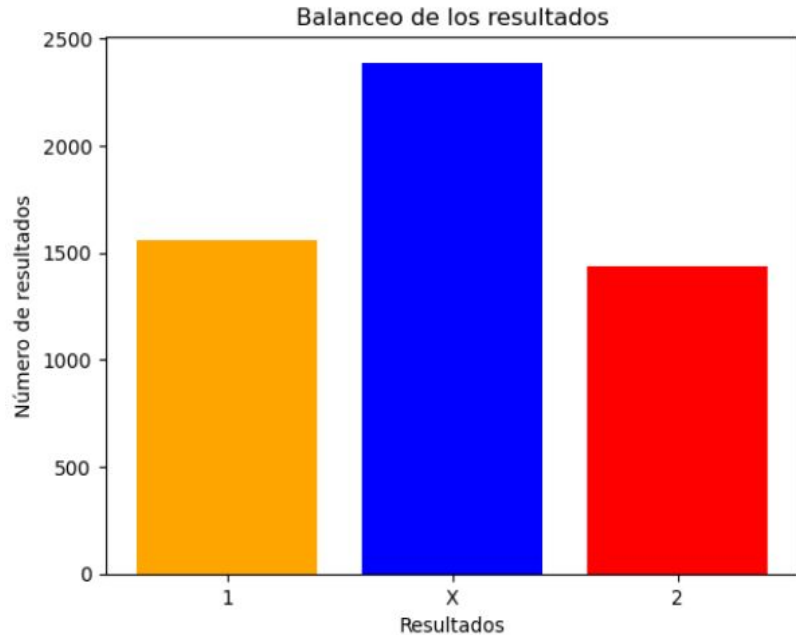
- La Quiniela: más de 10 aciertos.
- Apuestas (siempre con responsabilidad): apostando al ganador del encuentro.

OTROS POSIBLES USOS

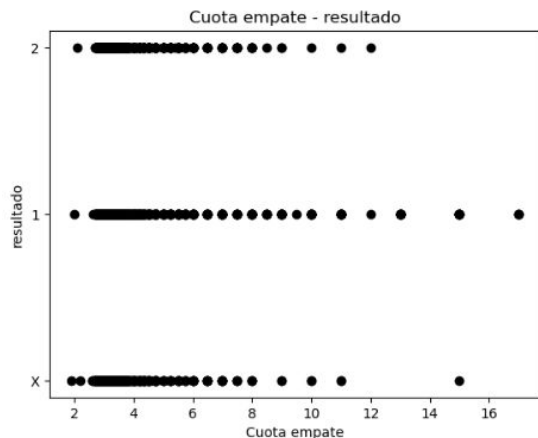
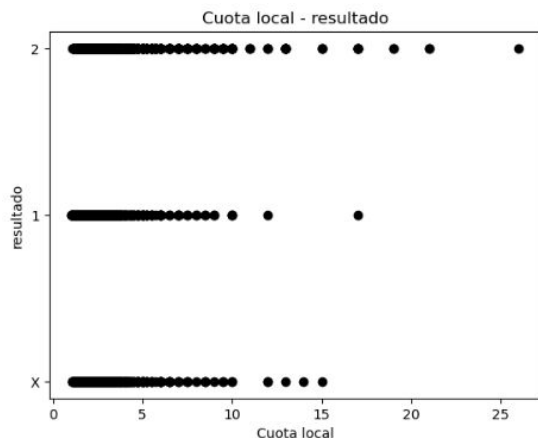
- Detección de cuotas mal colocadas.
- Regresión lineal para predecir en mercados como tarjetas, goles o corners.
- Escoger equipos correctamente (para evitar ser del Atleti).



PEQUEÑO ANÁLISIS EDA



PEQUEÑO ANÁLISIS EDA



Correlaciones con el target

SHOTS_ON_GOAL_AWAY → 0.250380

TOTAL_SHOTS_AWAY → 0.075284

SHOTS_INSIDEBOX_AWAY → 0.127434

RED_CARDS_AWAY → -0.051621

GOALKEEPER_SAVES_LOCAL → 0.058156

ODD_1 → 0.161175

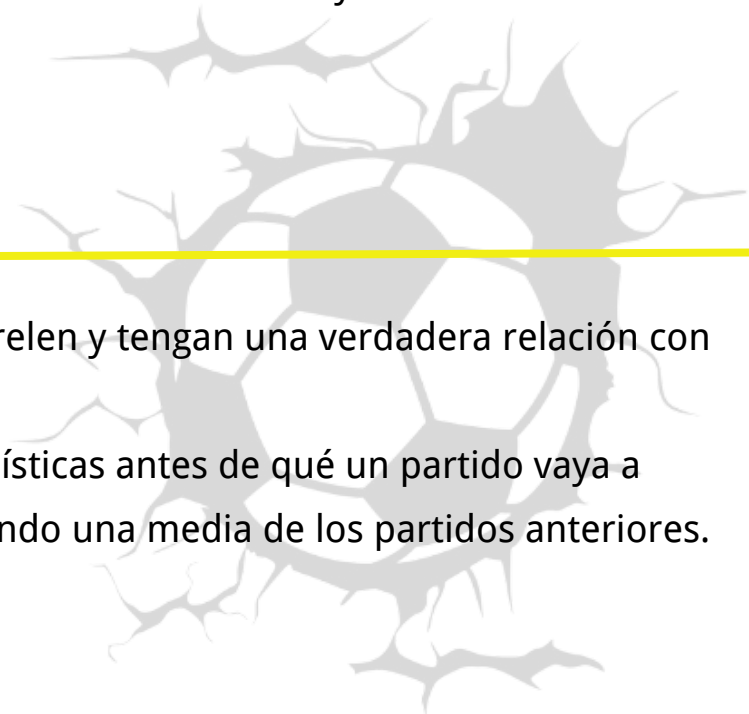
ODD_2 → -0.059482

MODELOS UTILIZADOS

Los modelos qué se han usado para clasificar los datos han sido RandomForest y XGBoost.

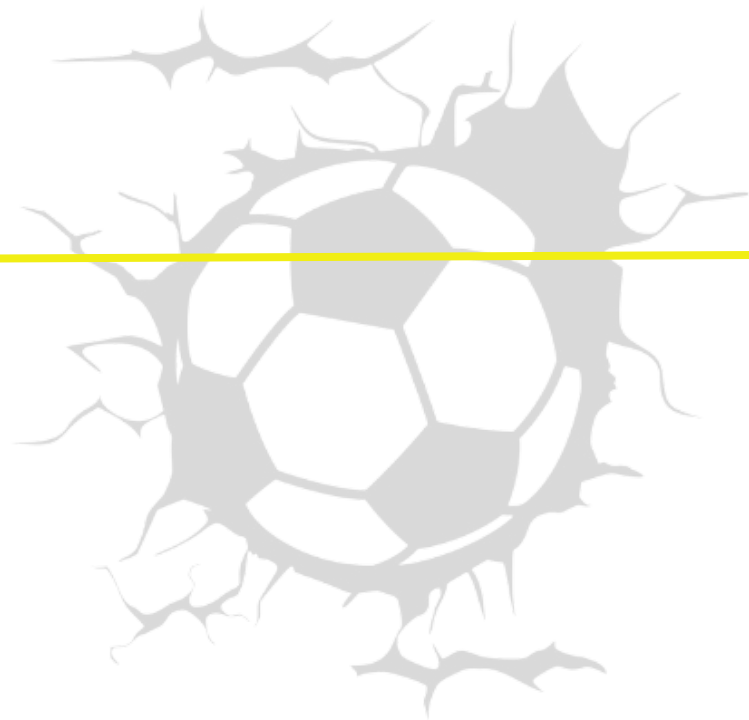
PROBLEMAS ENCONTRADOS

- No se han encontrado unas o varias variables qué corren y tengan una verdadera relación con el target, sin reflejarse tampoco en las estadísticas.
- A la hora de crear nuevos datos, no sabemos las estadísticas antes de qué un partido vaya a suceder. Por lo tanto, debemos asumir los datos, sacando una media de los partidos anteriores.



PREDICCIONES JORNADA 23.

PARTIDOS DE LA QUINIELA.



PARTIDOS

RANDOMFOREST

XGBOOST

I

X

2

Athletic Club			Girona	X	✗	2	✓	1.62	4	5.5	IO€	➡	+45€
Celta			Valladolid	1	✓	1	✓	1.67	4	5.25	IO€	➡	+6.7€
Cádiz			Rayo Vallecano	X	✗	X	✓	3	2.9	2.7	IO€	➡	+19€
Espanyol			Mallorca	1	✓	1	✓	2.6	2.8	3.25	IO€	➡	+16€
Almería			Barcelona	2	✗	2	✗	6.5	4.75	1.44	IO€	➡	-10€
Real Madrid			Atlético de Madrid	X	✓	X	✓	1.83	3.5	4.5	IO€	➡	+25€
Sevilla			Osasuna	X	✗	X	✗	1.7	3.6	5.25	IO€	➡	-10€
Villarreal			Getafe	X	✗	X	✓	1.73	3.5	5.75	IO€	➡	+25€
Valencia			Real Sociedad	X	✗	X	✗	3.2	3.1	2.38	IO€	➡	-10€
Alavés			Cartagena	1	✗	1	✓	1.7	3.25	6	IO€	➡	+7€
Sporting de Gijón			Tenerife	X	✗	X	✓	2.38	2.88	3.6	IO€	➡	+18.8€
Las Palmas			Ponferradina	1	✓	1	✓	1.65	3.6	5.5	IO€	➡	+6.5€
Ibiza			Huesca	X	✓	X	✗	3.6	2.8	3.1	IO€	➡	-10€
Mirandés			Leganés	1	✗	1	✗	2.9	2.9	2.7	IO€	➡	-10€
Zaragoza			Burgos	1	✗	1	✗	2.05	2.88	4.33	IO€	➡	-10€

INVERSIÓN
150€

BENEFICIO
169€

PERDIDO
60€

+109€

PREDICCIONES JORNADA 24.

PARTIDOS DE LA QUINIELA.



PARTIDOS

XGBOOST

XGBOOST (CON CUOTAS)

I

X

2

Getafe			Girona
Almería			Villarreal
Mallorca			Elche
Atlético de Madrid			Sevilla
Valladolid			Espanyol
Barcelona			Valencia
Rayo vallecano			Athletic Club
Betis			Real Madrid
Osasuna			Celta
Burgos			Granada
Huesca			Levante
Tenerife			Eibar
Leganés			Ibiza
Albacete			Sporting de Gijón
Lugo			Zaragoza

X	✗
2	✓
1	✗
X	✗
1	✓
1	✓
2	✗
2	✗
2	✗
1	✗
1	✗
1	✓
X	✗
2	✓
1	✓
1	✗

X	✗
2	✓
1	✗
X	✗
2	✗
1	✓
2	✗
2	✗
1	✗
1	✗
1	✓
1	✗
1	✓
X	✗
1	✗
1	✓
X	✓

2.62
3.4
1.65
1.6
2.25
1.33
2.75
4.75
2.6
2.75
3.75
2.3
1.91
2.05
3.6

10€
3.5
3.5
10€
3.8
3.25
5.25
3.1
3.8
3.1
2.8
3
10€
2.8
10€
2.9
3
10€
3

3	→	-10€
2.15	→	+24€
6	→	-10€
6	→	-10€
3.25	→	+15€
8.5	→	+3.3€
2.75	→	-10€
1.7	→	-10€
2.88	→	-10€
2.88	→	-10€
2.15	→	+27.5€
3.75	→	-10€
5.5	→	+45€
4.33	→	+10.5€
2.25	→	-10€

INVERSIÓN
150€

BENEFICIO
125.3€

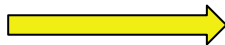
PERDIDO
90€

+35.3€

ELECCIÓN DEL MODELO

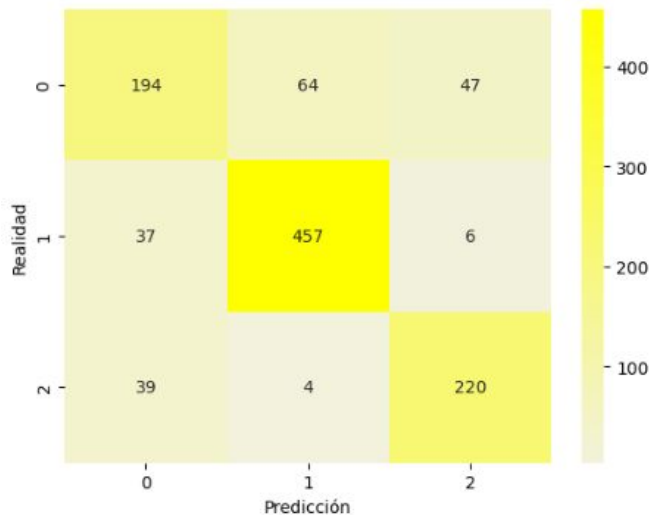
Por lo visto anteriormente, el modelo que utilizaremos para predecir jornada tras jornada, será el modelo de clasificación de XGBoost sin cuotas.

ACCURACY ENTRENANDO CON TRAIN Y PREDICIENDO CON TEST



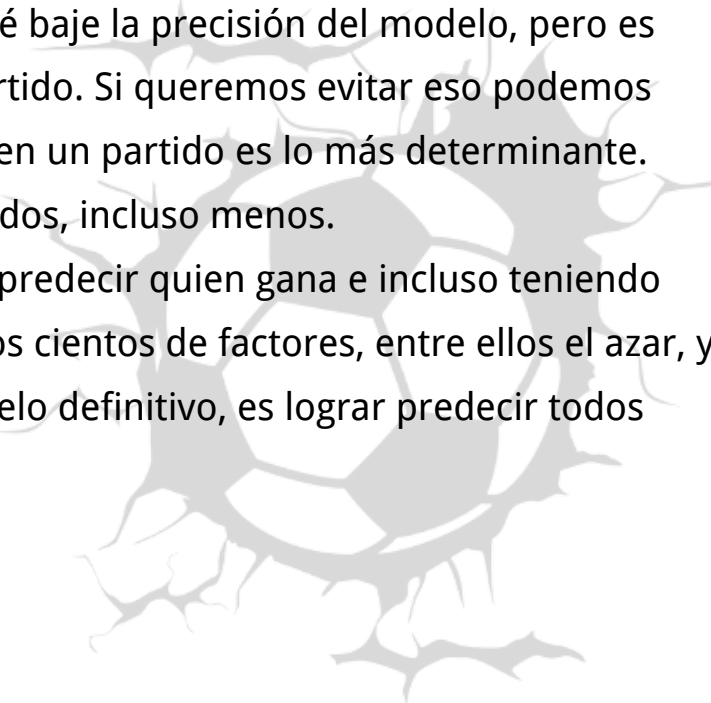
80.62%

50% REAL



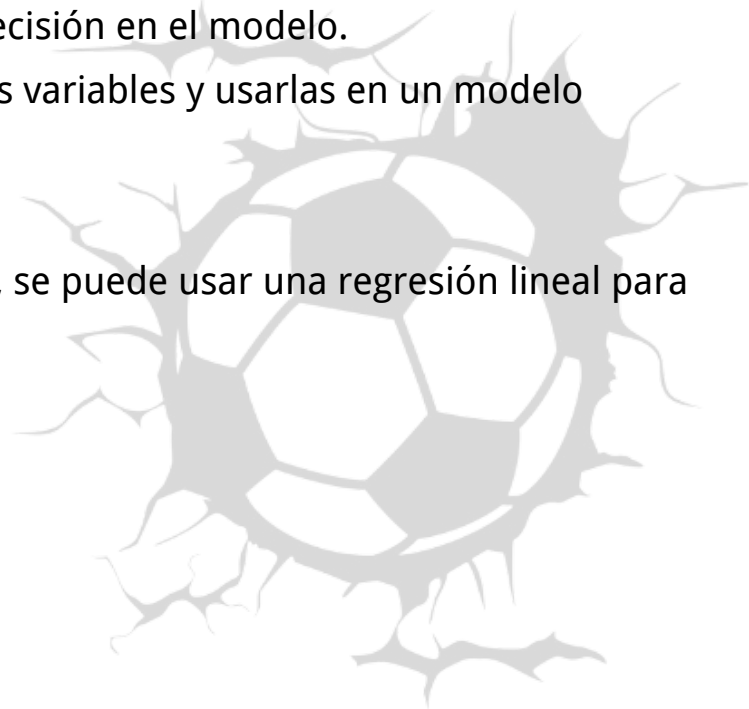
	Precision	Recall	F1-score	Support
0	0.72	0.64	0.68	333
1	0.87	0.91	0.89	500
2	0.81	0.84	0.82	263
accuracy			0.82	1068
macro avg	0.80	0.80	0.80	1068
weighted avg	0.81	0.82	0.81	1068

CONCLUSIONES

- Es muy complicado predecir el resultado de un partido de fútbol.
 - Predecir sin datos y teniendo que asumirlos, lleva a que baje la precisión del modelo, pero es imposible que sepamos qué va a pasar antes de un partido. Si queremos evitar eso podemos evitar las estadísticas como variable, pero lo que pasa en un partido es lo más determinante.
 - Se puede ganar dinero acertando el 50% de los resultados, incluso menos.
 - Por mucho que se pueda llegar a clasificar un partido, predecir quien gana e incluso teniendo sentido las probabilidades, en un partido influyen otros cientos de factores, entre ellos el azar, y puede ocurrir cualquier cosa. El objetivo final y el modelo definitivo, es lograr predecir todos ellos, para que el clasificador no falle.
- 

POSIBLES MEJORAS FUTURAS

- Búsqueda de variables que tengan una mayor relación con el resultado de un partido.
- Aplicar redes neuronales para obtener una mayor precisión en el modelo.
- Utilizar predicciones de varios modelos como posibles variables y usarlas en un modelo neuronal.
- Obtener datos de otras ligas.
- En lugar de utilizar la media de los últimos 3 partidos, se puede usar una regresión lineal para predecir las estadísticas de los partidos.





GRACIAS POR VUESTRA ATENCIÓN.

Proyecto Machine Learning

Gonzalo Velázquez Martínez

THE BRIDGE